15. Johnson LC, Letzel HW (1984) Measuring passive smoking: methods, problems, and perspectives. Prev Med 13:705–716
16. Koo LC, Ho JH-C, Saw D (1984) Is passive smoking an added risk factor for lung cancer in Chinese women? J Exp Clin Cancer Res 3:277–283
17. Lee PN, Chamberlain J, Alderson MR (1986) Relationship of passive smoking to risk of lung cancer and other smoking-associated diseases. Br J Cancer 54:97–105
18. Lehnert G (Chairman), Garfinkel L, Hirayama T, Schmahl D, Überla K, Wynder EL, Lee P (1984) Roundtable discussion. Prev Med 13:730–746
19. Lubin JH, Blot WJ (1984) Assessment of lung cancer risk factors by histologic category. JNCI 73:383–389
20. MacLennan R, Da Costa J, Day NE, Law CH, Ng YK Shanmugaratnam K (1977) Risk factors for lung cancer in Singapore Chinese, a population with high female incidence rates. Int J Cancer 20:854–860
21. Muramatsu M, Umemura S, Fukui J, Arai T, Kira S (1987) Estimation of personal exposure to ambient nicotine in daily environment. Int Arch Occup Environ Health 59:545–550
22. National Research Council, Committee on Passive Smoking, Board on Environmental Studies (1986) Environmental tobacco smoke. Measuring exposures and assessing health effects. National Academy Press, Washington, DC, pp 224–227
23. Pershagen G, Hrubec Z, Svensson C (1987) Passive smoking and lung cancer in Swedish women. Am J Epid 125:17–24
24. Rosner B (1982) Fundamentals of Biostatistics. Duxbury Press, Boston, MA, p 176
25. Sellers TA, Ooi WL, Elston RC, Chen VW, Bailey-Wilson JE, Rothschild H (1987) Increased familial risk for non-lung cancer among relatives of lung cancer patients. Am J Epid 126:237–246
26. Trichopoulos D, Kalandidi A, Sparros L, MacMahon B (1981) Lung cancer and passive smoking. Intl J Cancer 27:1–4
27. U.S. Department of Health and Human Services, Public Health Service (1982) The Health Consequences of Smoking: Cancer: a report of the Surgeon General. DHHS Publication No. (PHS)82-50179, p 243
28. U.S. Department of Health and Human Services. Public Health Service (1986) The health consequences of involuntary smoking. Report of the Surgeon General. DHHS Publication No. (CDC) 87-8398, pp 66–102
29. U.S. Department of Health, Education, and Welfare, Public Health Service (1964) Report of the Advisory Committee to the Surgeon General of the Public Health Service. Smoking and health. PHS Publication No. 1103, pp 174–175
30. Wu AH, Henderson BE, Pike MC, Yu MC (1985) Smoking and other risk factors for lung cancer in women. JNCI 74:747–751

Passive Smoking and Lung Cancer: A Reanalysis of Hirayama's Data

K. Überla and W. Ahlborn

The statistical association between environmental tobacco smoke and lung cancer is controversial. The Hirayama Study seems to provide sound epidemiological evidence supporting this hypothesis. In a recent paper [6] I have analyzed the published studies. Regarding the Hirayama study the following facts have to be kept in mind:

- The study was not designed to test the hypothesis, whether passive smoking is associated with lung cancer or not. It can therefore only generate this hypothesis, not prove it.
- The cohort was not representative for the population of Japan. A selection bias is possible.
- The exposure indicator – the fact of being married to a man who smokes – is not reliable, not valid and not specific.
- The event indicator – dying on lung cancer as noted on death certificates – is neither reliable nor valid.
- Various confounding factors – for instance exposure at the working place, indoor air pollution, overall air pollution, type of medical care – were not accounted for.
- Bias in registering the fact, that a woman is a nonsmoker, was not controlled. Resulting differential misclassifications of the cases, who were smokers and had to be excluded, have not been considered.
- Almost nothing is known about the 200 cases. No case reports are available, autopsy and histology are available in only 11.5%.

The core of the information, on which the results of this study rely, is

1) that during 1965 200 women in Japan told an interviewer on a single occasion that they were – during that time – nonsmokers and their husbands told that they were smokers, which might have been different before and afterwards and
2) that their death certificates subsequently contained the diagnosis lung cancer, which might have been erroneous.

Such sparse information does not seem to be convincing.

In our paper we consider three questions:

1) What is the relative risk when one removes the selection bias regarding age of women in the Hirayama cohort?
2) What is the relative risk for women married to men with different occupations, when one removes the selection bias regarding age of men?
3) What is the relative risk when additionally some differential misclassification is assumed?

H. Kasuga (Ed.) Indoor Air Quality

Material and Methods

We start from Tables 1, 2, and 3 of Hirayama 1984 [4]. These tables contain the most detailed published data. In order to check our program, we reproduced some of the reported relative risk estimates with good accuracy.

There are marked differences between the Hirayama cohort and the female age distribution over 40 in the population of Japan 1965. Women 50–59 are overrepresented, women older than 70 are severely underrepresented. In this age group only a single case from 12 was observed. The investigated cohort certainly has a severe selection bias by age, which needs no statistical test. This is likely due to the fact, that the smoking behaviour was not known in the elderly or that the husbands of older women have died. Since it takes 20 years and more from exposure to lung cancer, older women surely are relevant and should not be excluded. The majority of lung cancer cases occur in older age groups, in Germany more than 67% in women over 65 years.

In order to answer the question what the relative risk is when the age selection bias is removed, we adjusted the data to the age distribution of the female population of Japan.

Table 1. Differences between Hirayama cohort and the female age distribution over 40 in the population of Japan 1965*

Age group	Percent female	
	Japan population	Hirayama cohort
40–49	39	42
50–59	30	35
60–69	19	22
70 +	*12*	*1*
	100	100

* Population Census 1965. Statistical survey of economy of Japan; 1967. Ministry of Foreign Affairs of Japan.

Table 2. Smoking habit of husband by age of wife.* Original data

Wives age	Husbands smoking habit							
	Non		1–19		20 +		Total	
40–49	4	7,918	21	17,492	21	12,615	46	38,025
50–59	14	7,635	46	15,640	31	8,814	91	32,089
60–69	16	6,170	31	10,381	10	3,793	57	20,344
70 +	3	172	1	671	2	239	6	1,082
Total	37	*21,895*	99	*44,184*	64	*25,461*	200	*91,540*

* Table 2 of Hirayama 1984.

Table 3. Smoking habit of husband by age of wife*. Removed selection bias: Data adjusted to the age distribution of women in the population

Wives age	Husbands smoking habit							
	Non		1–19		20 +		Total	
40–49	3.91	7,748.8	19.12	15,927.8	20.02	12,024.0	43.05	35,700.6
50–59	12.49	6,813.7	38.20	12,987.1	26.95	7,661.2	77.64	27,462.0
60–69	14.25	5,496.6	25.70	8,604.9	8.68	3,291.1	48.63	17,392.6
70 +	32.02	1,835.9	9.93	6,664.2	20.79	2,484.7	62.74	10,984.8
Total	62.67	*21,895*	92.95	*44,184*	76.44	*25,461*	232.06	*91,540*

* Table 2 of Hirayama 1984.

The technique of iterative proportional fitting of a contingency table to given marginals as described by Bishop et al. [1] or by Hartung et al. [3] was used. This technique keeps the risks constant as observed in every cell and changes the marginals and the cell counts according to the given age distribution of the population. Iterative proportional fitting of contingency tables to given marginals is a well known technique in multivariate statistics and can be applied here without changing the observed interrelations between smoking habit, occupation, and lung cancer. From the fitted or adjusted tables the risk ratios are calculated in the usual way. Such risk ratios based on data with removed age selection bias are the correct ones and should be used.

One has to require that there should be no selection bias by age and the cases should be included as they would have occured in the population. Otherwise statistical tests and p-values are not very meaningful.

Table 2 shows the original data by age of wife. The cells contain the number of lung cancer cases and those under risk as published by Hirayama. The 1–19 group includes ex-smokers in this and the following tables. 200 cases out of 91,540 women were observed. Iterative proportional fitting to the female age distribution of the population leaves the hatched numbers constant. The others are adjusted using a right hand marginal which is made proportional to the age distribution of the population.

Results

Table 3 gives the results of iterative proportional fitting to the female age distribution of the population. It contains the numbers of those under risk and of lung cancer deaths as they would have been observed, if Hirayama had not excluded or preferred certain age groups. The age selection bias is removed. The risks in the individual cells are still the same as those observed by Hirayama. Also the structure of the common distribution regarding age, smoking habit and lung cancer is unchanged. Hirayama would have totally observed 232 cases instead of 200, with the corresponding numbers in the individual cells, had he included all women as they live in the population. This table is the best available starting point for age-standardized risk ratio calculations. It was not used so far.

Table 4. Relative risk by age of women*

	Husbands smoking habit		
	Non	1–19	20 +
$\widehat{RR}$	1.00	*1.37*	*1.56*
IL_{90}		1.00	1.11
MH-CHI		1.51	2.27
$P_{one\ tailed}$		0.065	0.012**
$\widehat{RR}$	1.00	*0.77*	*1.06*
IL_{90}		0.59	0.80
MH-CHI		2.19	0.27
$P_{one\ tailed}$		0.014***	0.395

Upper part: standardized by age of women only.
Lower part: age selection bias removed and standardized by age of women.
$\widehat{RR}$: Weighted point estimate of rate ratio.
IL_{90}: Lower 90-percent confidence interval.
* Calculated from Table 2 of Hirayama 1984.
** "Significant" in positive direction.
*** "Significant" in negative direction.

In the upper part of Table 4 you find the risk ratios standardized by age only, as done by Hirayama. The lower part are the risk ratios after removing the age selection bias. In the upper part the weighted point estimate of the rate ratio is 1.56 in the 20+-group and is technically "significant". IL_{90} designates the lower point of the 90-percent confidence interval in this and the following tables, as it was used by Hirayama.

This risk increase disappears completely when one removes the selection bias by age. In the 20+-group the rate ratio is 1.06, hardly a relevant risk increase. In the group of 1–19 cigarettes per day it is 0.77 which is a technically significant risk decrease. The adjusted rate ratio, considering all those exposed in one group versus those not exposed is 0.901 with a confidence interval including unity. If Hirayama had observed the cases as they occur in the female population without selection bias by age, he would have observed no risk increase, but a risk decrease. This is the main result of our reanalysis, which corresponds well with the result of the prospective American cohort study as published by Garfinkel [2].

We now consider two occupations, farmers and industry workers. From the upper part of Table 5 one can see that the relative risk for wives of farmers seems substantial, when one standardizes by age of men only. The point estimates of the rate ratios are 1.48 and 1.63 respectively. This was observed earlier and had no adequate explanation. If one removes the selection bias by age and adjusts to the male age distribution of Japan – the numbers in the lower part of Table 5 – the rate ratios are 0.85 and 0.82, not different from unity. This seems more plausible.

Considering the wives of industry workers only, in the upper part of Table 6, the point estimates of the rate ratios are 1.77 and 2.27, standardized by age of men, being not significant. Removing the age selection bias – in the lower part of Table 6 – there is a remarkable risk increase to 4.60 and 6.90, which is significant. However, there are only

Table 5. Relative risk: wives of farmers only*

	Husbands smoking habit		
	Non	1–19	20 +
$\widehat{RR}$	1.00	*1.48*	*1.63*
IL_{90}		0.97	1.01
MH-CHI		1.48	1.92
$P_{one\ tailed}$		0.069	0.027
$\widehat{RR}$	1.00	*0.85*	*0.82*
IL_{90}		0.59	0.53
MH-CHI		0.42	0.53
$P_{one\ tailed}$		0.337	0.296

Upper part: standardized by age of men only.
Lower part: age selection bias removed and standardized by age of men.
$\widehat{RR}$: Weighted point estimate of rate ratio.
IL_{90}: Lower 90-percent confidence interval.
* Calculated from Table 3 of Hirayama 1984.

Table 6. Relative risk: wives of industry workers only*

	Husbands smoking habit		
	Non	1–19	20 +
$\widehat{RR}$	1.00	*1.77*	*2.27*
IL_{90}		0.70	0.84
MH-CHI		0.73	0.81
$P_{one\ tailed}$		0.232	0.208
$\widehat{RR}$	1.00	*4.60*	*6.90*
IL_{90}		1.71	2.45
MH-CHI		2.50	2.78
$P_{one\ tailed}$		0.006	0.003

Upper part: standardized by age of men only.
Lower part: age selection bias removed and standardized by age of men.
$\widehat{RR}$: Weighted point estimate of rate ratio.
IL_{90}: Lower 90-percent confidence interval.
* Calculated from Table 3 of Hirayama 1984.

9 lung cancer deaths in the 20+-group and only 3 in women 70 years and older, which are small numbers, but these are numbers observed and used by Hirayama and his risk structure is unchanged. Thus only in the subgroup of women married to industry workers there is a risk increase, in all other occupations there is no risk increase. Omitting industry

Table 7. Relative risk: assumed differential misclassifications*

Number of cases assumed misclassified and removed from exposed groups		Husbands smoking habit		
		Non	1–19	20 +
n = 10 = 5%	$\widehat{RR}$	1.00	*0.74*	*1.00*
	$P_{one\ tailed}$		0.006	0.469
n = 20 = 10%	$\widehat{RR}$	1.00	*0.70*	*0.93*
	$P_{one\ tailed}$		0.003	0.383
n = 30 = 15%	$\widehat{RR}$	1.00	*0.66*	*0.85*
	$P_{one\ tailed}$		0.001	0.238

Age selection bias removed and standardized by age of women.

$\widehat{RR}$: Weighted point estimate of rate ratio.

* Calculated from Table 2 of Hirayama 1984.

workers, the point estimates of the rate ratios are 0.90 and 0.89, not significantly different from unity. These findings are consistent with the assumption of confounding factors in women married to industry workers, who might be exposed to other environmental hazards. Our calculations show that by removing selection bias by age, one can explain hitherto implausible results.

Active smoking is correlated among married couples. In a society in which female smokers were very rare in 1965, more women married to smokers will declare themselves nonsmokers than the other way round. One has therefore to considerr biased or differential misclassification. There are likely more women with lung cancer, who have been misclassified as nonsmokers and have to be removed from the cohort, than the other way round.

We made some moderate assumptions regarding differential misclassification, as shown in Table 7. In order to examine how sensitive the relative risk is we removed 10, 20, and 30 cases from the exposed groups – corresponding to 5, 10, and 15 percent.

Assuming 30 misclassified cases – 15 percent, a percentage which has been observed in the literature [5] – the rate ratios are 0.66 and 0.85. In the group 1–19 cigarettes per day all the risk estimators are significantly smaller than unity. Our personal opinion is that 10 differential misclassified cases from 200, who have to be omitted, are a fair number. The corresponding weighted point estimates of the rate ratio are 0.74 and 1.00. These risk estimates are as reasonable as other risk estimates calculated from the Hirayama data. They indicate – if anything – a risk decrease, not a risk increase.

Discussion

Reanalyses of data, which have been collected by others are not easy. This is because information is not completely available, because information might be misinterpreted or because one has to take another view in order to come closer to the acceptable truth. Our calculations do not diminish the great value and impact the Hirayama study had on the epidemiology of passive smoking. They show however, that reasonable alternative views

Table 8. Reanalysis of Hirayama's data: summary of relative risk

		Husbands smoking habit		
		Non	1–19	20 +
Age selection bias removed and age-standardized (women)	$\widehat{RR}$	1.00	*0.77*	*1.06*
	$P_{one\ tailed}$		0.014	0.395
Without industry workers, age selection bias removed and age-standardized (men)	$\widehat{RR}$	1.00	*0.90*	*0.89*
	$P_{one\ tailed}$		0.394	0.179
10 cases assumed misclassified, age selection bias removed and age-standardized (women)	$\widehat{RR}$	1.00	*0.74*	*1.00*
	$P_{one\ tailed}$		0.006	0.469

$\widehat{RR}$: Weighted point estimate of rate ratio.

on the same data are possible, which lead to opposite conclusions. Our findings are in contrast to Hirayama's thesis that – based on his data – there is a substantial statistical association between passive smoking and lung cancer.

As long as there is no other independent and sound epidemiological evidence, it should be left to the individual scientist which analysis of the same data he thinks is more appropriate. We do not hold that our view is the only correct one. We do hold however, that the risk ratios calculated by us, removing age selection bias, are as valid as other risk estimates. To our opinion they are more appropriate, since they go back to the population and not to a selected sample. Even when one would take another marginal, for instance the age distribution of wives still married to living men – which was not available – the effect would be considerable. Our risk estimates are a consequence of the data published by Hirayama and cannot be rejected from the study data, as they are published so far.

To summarize (Table 8): Removing the age selection bias in the Hirayama study one gets a relative risk of 1.06 in the group of women married to men with more than 20 cigarettes per day. In the group of women married to men with 1–19 cigarettes per day the relative risk is 0.77, a technically "significant" risk decrease. If Hirayama could have observed the lung cancer cases as they occur in the female population, he would have observed no risk increase, but a risk decrease to around 0.90, considering those exposed versus those not exposed. This fact deserves attention.

If one omits the wives married to industry workers because of possible confounding factors in this group, the relative risk is 0.90 and 0.89 respectively. This is of the same size order and smaller than unity. Here we could adjust and standardize by occupation and age of men only, which is not as appropriate as by the age of women.

If one assumes that 10 cases are differentially misclassified and removes them from the exposed groups, the risk estimates are 0.74 and 1.00, respectively. Our findings demonstrate how sensitive the data of this study are and how weak the evidence for a statistical association between passive smoking and lung cancer might be. In view of these and other facts some of which we mentioned in the introduction, the null hypothesis might be true as well and seems to be consistent with the Hirayama data in the same way as the alternative hypothesis.

We would be glad to apply our technique to more detailed data if we can get them from Hirayama, for instance in order to adjust by occupation of men and age of women, or by occupation of men and by age of women married to a husband who is still alive. We are ready to modify our view if such data can support the alternative hypothesis better than the published data. We do hope, that our calculations give rise to a fruitful discussion. The methods we used here might be of interest to the analysis of other cohort and case control studies.

References

1. Bishop YV, Fienberg StE, Holland PW (1980) Discrete multivariate analysis: theory and practice. MIT Press, Cambridge, pp 97
2. Garfinkel L (1984) Time trends in lung cancer mortality among nonsmokers and a note on passive smoking. J Natl Canc Inst (USA) 66:1061–1066
3. Hartung J, Elpelt B, Klösener KH (1985) Statistik. Lehr- und Handbuch der angewandten Statistik. München Wien Oldenburg, p 501–503
4. Hirayama T (1984) Lung cancer in Japan: effects of nutrition and passive smoking. In: Mizell M, Correa P (eds) Lung cancer: causes and prevention. Chemie, Weinheim, pp 175–195
5. Lee PN, Chamberlain J, Alderson HR (1986) Relationship of passive smoking to risk of lung cancer and other smoking-associated diseases. Br J Cancer (54):97–105
6. Überla K (1987) Lung cancer from passive smoking: hypothesis or convincing evidence? Int Arch Occup Environ Health (59):421–431

What Is the Epidemiologic Evidence for a Passive Smoking-Lung Cancer Association?

N. Mantel

Summary

Two survey articles of reports on the association of passive smoking with lung cancer have recently appeared, and also a comprehensive report on the subject of environmental tobacco smoke by a committee of the National Research Council of the United States. The observed excess over a relative risk of unity cannot be explained by chance. Nor can it be fully accounted for by a particular source of bias, the false claims of being non-smokers by individuals who were active or ex-smokers. That possible source of bias leads, in one summary survey, to reducing a relative risk of 1.35 to 1.30, but from 1.34 to 1.15 in the National Research Council report. The latter report suggests that statistical significance would no longer obtain, perhaps, particularly, because of other possible biases. However, to get an estimate of the correct relative risk due to passive smoking, allowance has to be made for actual exposure to passive smoking of those not exposed at home. Thus, the 1.30 is adjusted upwards, by 18 in one survey, to 1.53, but by only 8% in the National Research Council report to 1.24. The National Research Council report had given an anticipated relative risk of 1.1 based on dosimetric considerations. But it is suggested here that that could be as low as 1.05, too low to be detected in an epidemiologic investigation – in any case it would be based on hypothetical assumptions.

In November of 1986 there were two near-simultaneous review articles addressing the subject of passive smoking and lung cancer. One was an invited guest editorial by Blot and Fraumeni in the Journal of the National Cancer Institute, the other a contemporary theme discussion by Wald et al. in the British Medical Journal [1, 2].

There was substantial overlapping in the two articles of the various publications on the subject, and on the basis of which the conclusion of a significant positive association was made. The article by Wald et al. gave, perhaps, more statistical detail about the results of the several studies covered. But, to my mind, there was uncritical acceptance of the results of all the studies. Blot and Fraumeni did suggest that there were some flaws in a particular study, that by Hirayama [3], but decided that any inherent biases in that investigation could not have given rise to the observed elevated risk.

From their overall evaluation of 10 case-control studies (all 10 gave results for females, five separately for males as well) and three prospective studies (two of these covered males separately), which provided 20 separate relative risk (actually odds ratio) values, Wald et al. came up with a summary relative risk of lung cancer due to passive smoking of 1.35 (95% limits 1.19 to 1.54). They trim this down to 1.30 on the basis that some of the presumed non-smokers exposed to passive smoking were actually smokers. Then, on the added basis that even those unexposed to passive smoking at home may still have been exposed when away from home, they raise their estimate of relative risk to 1.53. But note that this last modification presupposes the answer, that passive smoking does

H. Kasuga (Ed.) Indoor Air Quality